

Qy 169 ASKIFNCRMWEEKVERGRRTSLCTHDPAKICSRDH-----AQSSATWSCSQP 215  
 ||| | : : ||: || || | | :| :||:|  
 Db 1358 CKKIFAC-----KYKECNKRFLCSKALAKHCSDSHNLDHIEEPKVLSEAGSAARFSCNQP 1412

Qy 216 FKVVCVYIAFYSTDYRLVQ-----KVC PDYNYHSD 245  
 : | : | : | : : || | |  
 Db 1413 -QCPAVFYTFNKLKHHLM EQHNIEGEIHSDYEIHCD 1447

Search completed: June 14, 2004, 19:08:10  
 Job time : 29 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2004, 18:59:26 ; Search time 20 Seconds  
 (without alignments)  
 1212.013 Million cell updates/sec

Perfect score: 1386  
 Sequence: 1 MQLTRCCFVFLVQGS LYLVI.....VQKVC PDYNYHSDTPYPSG 252

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : PIR\_78:\*  
 1: pir1:\*  
 2: pir2:\*  
 3: pir3:\*  
 4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	% Query		Match Length	DB ID	Description
	Score	Match			
1	97.5	7.0	264	1 S35219	homeotic protein H
2	96	6.9	1838	1 CGHU1V	collagen alpha 1(V

3	94	6.8	1843	2	S18803	collagen alpha 1(V
4	92.5	6.7	669	1	VCVWEK	env polyprotein -
5	91	6.6	117	2	D40750	proline-rich prote
6	90.5	6.5	264	1	WJHU3E	homeotic protein H
7	90.5	6.5	858	2	JG0183	myosin Myok - Dict
8	90	6.5	1560	2	T02885	peroxisome prolife
9	89.5	6.5	669	2	A46511	envelope protein -
10	89.5	6.5	750	2	D86245	hypothetical prote
11	89.5	6.5	1001	2	S30385	G9a protein - huma
12	89	6.4	2090	2	S26058	probable transform
13	88.5	6.4	574	2	T04249	hypothetical prote
14	88.5	6.4	1175	2	I57549	adenosine deaminas
15	87.5	6.3	180	2	S43791	PBDX protein - hum
16	87	6.3	1690	1	CGHU1B	collagen alpha 4(I
17	86.5	6.2	401	2	S75288	sensory transducti
18	86.5	6.2	1603	2	S23810	collagen alpha 1(X
19	86	6.2	392	1	PIHUB6	salivary proline-r
20	86	6.2	644	2	S15464	gp70 protein - mur
21	85.5	6.2	1712	1	CGHU2B	collagen alpha 2(I
22	85	6.1	610	2	T22540	hypothetical prote
23	84.5	6.1	407	2	S22586	homeotic protein E
24	84.5	6.1	463	1	EFHUA2	translation elonga
25	84.5	6.1	463	2	A40389	translation elonga
26	84.5	6.1	463	2	JC2445	translation elonga
27	84.5	6.1	609	1	A43458	replication protei
28	84.5	6.1	743	1	S23779	collagen alpha 1(V
29	84.5	6.1	1300	2	T03166	probable immediate
30	84	6.1	689	2	B43491	env polyprotein -
31	84	6.1	1027	2	T27970	hypothetical prote
32	84	6.1	1870	2	S37671	MHC class III hist
33	84	6.1	1872	2	S36152	MHC class III hist
34	83.5	6.0	226	2	JH0231	thaumatin-like pro
35	83	6.0	688	2	A43491	env polyprotein -
36	83	6.0	2142	2	B35098	MHC class III hist
37	82.5	6.0	188	2	JH0481	basic proline-rich
38	82.5	6.0	315	2	T47181	hypothetical prote
39	82.5	6.0	443	2	S21909	translation elonga
40	82.5	6.0	714	2	S77385	nitrate reductase
41	82	5.9	235	1	Q TTC2	thaumatin II precu
42	82	5.9	295	2	A44984	collagen - nematod
43	82	5.9	443	1	I38239	transcription fact
44	82	5.9	638	2	JC2139	phenylethylamine o
45	82	5.9	1400	1	I38185	protein-tyrosine k

#### ALIGNMENTS

##### RESULT 1

S35219

homeotic protein Hox C4 - mouse

N;Alternate names: homeotic protein Hox 3.5; homeotic protein MAB87

C;Species: Mus musculus (house mouse)

C;Date: 10-Dec-1993 #sequence\_revision 03-Aug-1995 #text\_change 22-Jun-1999

C;Accession: S35219; A49153; C41606; I49752

R;Goto, J.; Miyabayashi, T.; Wakamatsu, Y.; Takahashi, N.; Muramatsu, M.

Mol. Gen. Genet. 239, 41-48, 1993

A;Title: Organization and expression of mouse Hox3 cluster genes.  
 A;Reference number: S35219; MUID:93288004; PMID:8099712  
 A;Accession: S35219  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-264 <GOT>  
 A;Cross-references: GB:S62287; NID:g385749; PIDN:AAB27153.1; PID:g385750;  
 GB:D11328; NID:g406212; PID:g416420  
 A;Note: entry MUSHOX35A in GenBank release 103 duplicates GenBank entry S62287  
 except for an incorrect reference citation  
 R;Geada, A.M.; Gaunt, S.J.; Azzawi, M.; Shimeld, S.M.; Pearce, J.; Sharpe, P.T.  
 Development 116, 497-506, 1992  
 A;Title: Sequence and embryonic expression of the murine Hox-3.5 gene.  
 A;Reference number: A49153; MUID:93161956; PMID:1363091  
 A;Accession: A49153  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-79,'G',81-95,'S',97-264 <GEA>  
 A;Cross-references: GB:X69019; NID:g396183; PIDN:CAA48784.1; PID:g396184  
 A;Note: sequence extracted from NCBI backbone (NCBIN:124829, NCBIP:124830)  
 R;Murtha, M.T.; Leckman, J.F.; Ruddle, F.H.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 10711-10715, 1991  
 A;Title: Detection of homeobox genes in development and evolution.  
 A;Reference number: A41606; MUID:92073357; PMID:1720547  
 A;Accession: C41606  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 177-201 <MUR>  
 A;Cross-references: GB:M81660; NID:g193975; PIDN:AAA63313.1; PID:g193976  
 C;Genetics:  
 A;Gene: Hoxc-4; Hox 3.5  
 A;Map position: 15  
 A;Introns: 147/1  
 C;Function:  
 A;Description: control of embryonic development by tissue- and stage-specific  
 regulation of transcription  
 C;Superfamily: homeotic protein Hox D4; homeobox homology  
 C;Keywords: DNA binding; embryo; homeobox; nucleus; transcription regulation  
 F;157-213/Domain: homeobox homology <HOX>

Query Match 7.0%; Score 97.5; DB 1; Length 264;  
 Best Local Similarity 22.6%; Pred. No. 0.55;  
 Matches 48; Conservative 25; Mismatches 90; Indels 49; Gaps 9;

Qy	27	PPGSEDPERD-----DHEGQPRPRVPRKRGHISP-KSRPMAN-STLLGLLAPPGEAWGI	78
		:        :  :	
Db	57	PPRPSYPERQYSCTSLQPGNSRAHGPAQAGHHHPEKSQPLCEPAPLSGTSASPSAPPA	116
Qy	79	LGQPPNRPNHSPPPSAKVKKIFGWGDFYSNIKTVALNLLVTGKIVDHGNGTFS-----	131
		:  :  :    :    :	
Db	117	CSQP--APDHPSSAASKQPIVYPW---MKKIHVSTVNPNYNGGEPKRSRTAYTRQQVLEL	171
Qy	132	---VHFQHNATGQGNISISLVPPSKAVEFHQEQQIFIEAKASKIF--NCRMWEKVERGR	186
		:   :     : : :       :	
Db	172	EKEFHYNRYLTRRRRIEIA-----HSLCLSERQIKIWFQNRMRMKWKDHRLP	218
Qy	187	RTSLCTHDPKIC-----SRDHAQSS	207

Db                   | : : ||                   | ||:|:|:  
219 NTKVRSAPPAGAAPSTLSAATPGTSEDHSQSA 250

RESULT 2

CGHU1V

collagen alpha 1(V) chain precursor - human

N;Alternate names: procollagen alpha 1(V) chain

C;Species: Homo sapiens (man)

C;Date: 22-Nov-1993 #sequence\_revision 03-Oct-1995 #text\_change 16-Jun-2000

C;Accession: S18802; S16024; A61142; S11303; S03978; S43642; S58665

R;Greenspan, D.S.; Cheng, W.; Hoffman, G.G.

J. Biol. Chem. 266, 24727-24733, 1991

A;Title: The pro-alpha1(V) collagen chain. Complete primary structure, distribution of expression, and comparison with the pro-alpha1(XI) collagen chain.

A;Reference number: S18802; MUID:92105142; PMID:1722213

A;Accession: S18802

A;Molecule type: mRNA

A;Residues: 1-1838 <GRE>

A;Cross-references: GB:M76729; NID:g189519; PIDN:AAA59993.1; PID:g189520

R;Takahara, K.; Sato, Y.; Okazawa, K.; Okamoto, N.; Noda, A.; Yaoi, Y.; Kato, I.

J. Biol. Chem. 266, 13124-13129, 1991

A;Title: Complete primary structure of human collagen alpha-1(V) chain.

A;Reference number: S16024; MUID:91302336; PMID:2071595

A;Accession: S16024

A;Molecule type: mRNA

A;Residues: 1-81,'QL',84-389,'A',391-676,'K',678-1294,'PS',1297,'RS',1300-1553,'R',1555-1812,'V',1814-1838 <TAK>

A;Cross-references: GB:D90279; NID:g219509; PIDN:BAA14323.1; PID:g219510

A;Note: parts of this sequence were determined by protein sequencing

R;Yaoi, Y.; Hashimoto, K.; Takahara, K.; Kato, I.

Exp. Cell Res. 194, 180-185, 1991

A;Title: Insulin binds to type V collagen with retention of mitogenic activity.

A;Reference number: A61142; MUID:91224163; PMID:1709100

A;Accession: A61142

A;Molecule type: protein

A;Residues: 823-824,'X',826-842 <YAO>

A;Note: the residue designated 'X' is probably glycosylated hydroxylysine; this cyanogen bromide fragment contains an uncharacterized growth hormone-binding region

R;Yaoi, Y.; Hashimoto, K.; Koitabashi, H.; Takahara, K.; Ito, M.; Kato, I.

Biochim. Biophys. Acta 1035, 139-145, 1990

A;Title: Primary structure of the heparin-binding site of type V collagen.

A;Reference number: S11303; MUID:90366601; PMID:2203476

A;Accession: S11303

A;Molecule type: protein

A;Residues: 823-824,'X',826-848,'I',850-851,'P',853,'PR',856-893,'D',895-932,'X',934-950 <YA2>

A;Note: the residues designated 'X' are probably glycosylated hydroxylysine; this sequence revised by S16024

R;Seyer, J.M.; Kang, A.H.

Arch. Biochem. Biophys. 271, 120-129, 1989

A;Title: Covalent structure of collagen: amino acid sequence of three cyanogen bromide-derived peptides from human alpha-1(V) collagen chain.

A;Reference number: S03978; MUID:89227189; PMID:2496661

A;Accession: S03978

A;Molecule type: protein  
A;Residues: 621-640,'G',642-649,'L',651-662,'E',664-667,'Q',669-676,'Q',678-683,'P',685-691,'VT',694,'E',696-697,'AP',700-726,'Q',728-740,'L',742-746,'Q',748-752,'A',754-758,'N',760-775,'QK',778-822 <SEY>  
A;Note: there are a number of inconsistencies between the sequences in figures 6 and 7; the sequence in figure 7 is given  
R;Moradi-Ameli, M.; Rousseau, J.C.; Kleman, J.P.; Champliand, M.F.; Boutillon, M.M.; Bernillon, J.; Wallach, J.; van der Rest, M.  
Eur. J. Biochem. 221, 987-995, 1994  
A;Title: Diversity in the processing events at the N-terminus of type-V collagen.  
A;Reference number: S43642; MUID:94237164; PMID:8181482  
A;Accession: S43642  
A;Molecule type: protein  
A;Residues: 565-576;756-758,'X',760-763,'X',765-772;1012-1029;1219-1232;1465-1474,'X',1476-1477 <MOR>  
R;Fessler, L.I.; Brosh, S.; Chapin, S.; Fessler, J.H.  
J. Biol. Chem. 261, 5034-5040, 1986  
A;Title: Tyrosine sulfation in precursors of collagen V.  
A;Reference number: A56977; MUID:86168226; PMID:3082875  
A;Contents: annotation; identification of tyrosine sulfate in the amino-terminal propeptide  
R;Lee, S.; Greenspan, D.S.  
Biochem. J. 310, 15-22, 1995  
A;Title: Transcriptional promoter of the human alpha-1(V) collagen gene (COL5A1).  
A;Reference number: S58665; MUID:95374437; PMID:7646438  
A;Accession: S58665  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-36 <LEE>  
A;Cross-references: GB:L38808; NID:g1020325; PIDN:AAA79853.1; PID:g1020326  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated to varying extents. Prolines are predominately 4-hydroxylated. About 50% of the lysines are 5-hydroxylated and subsequently O-glycosylated.  
C;Comment: A long form of the mature protein containing part of the amino-terminal propeptide has been detected but not characterized. The homotrimer is probably fully processed to the short form, while the heterotrimers are probably processed to the long form.  
C;Genetics:  
A;Gene: GDB:COL5A1  
A;Cross-references: GDB:131457; OMIM:120215  
A;Map position: 9q34.2-9q34.3  
C;Complex: type V collagen may be a homotrimer of alpha 1(V) chains, a heterotrimer of two alpha 1(V) chains and one alpha 2(V) chain (see PIR:CGHU2V), or a heterotrimer of one alpha 1(V) chain, one alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among their carboxyl-terminal propeptides; a polymer of collagen trimers, offset by approximately one-quarter of their length, is formed with desmosine cross-links made from lysine and allysine residues  
C;Function:  
A;Description: structural component of extracellular fibrous polymer associated with cell surfaces and interstitial fibrils; widely distributed but least abundant of the fibrillar collagens  
A;Note: may play a role in controlling the lateral growth of collagen I fibrils

C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology  
 C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline; pyroglutamic acid; sulfoprotein; trimer; triple helix  
 F;1-37/Domain: signal sequence #status predicted <SIG>  
 F;36-261/Domain: PARP-like #status predicted <PARP>  
 F;38-541/Domain: amino-terminal propeptide #status predicted <PRO>  
 F;542-1605/Product: collagen alpha 1(V) chain, short form #status predicted <MAT>  
 F;542-558/Region: amino-terminal nonhelical telopeptide  
 F;559-1572/Region: helical  
 F;645-647/Region: cell attachment (R-G-D) motif  
 F;663-665/Region: cell attachment (R-G-D) motif  
 F;897-929/Region: heparin binding  
 F;1573-1605/Region: carboxyl-terminal nonhelical telopeptide  
 F;1606-1838/Domain: carboxyl-terminal propeptide #status predicted <CPR>  
 F;1615-1837/Domain: fibrillar collagen carboxyl-terminal homology <FCC>  
 F;38/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F;62-244,183-237/Disulfide bonds: #status predicted  
 F;159,176,385,1672,1741/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F;234,236,240,262,263,273,274,275,277,279,280,338,340,346,347,352,357,416,417,420,421/Binding site: sulfate (Tyr) (covalent) #status predicted  
 F;535/Modified site: allysine (Lys) #status predicted  
 F;541-542/Cleavage site: Ala-Gln (procollagen N-endopeptidase) #status predicted  
 F;542/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F;570,576,621,639,648,654,657,675,678,690,693,696,705,717,720,726,732,741,750,753,756,762,765,771,780,789,816,834,870,873,876,888,891,903,906,930,945,1017,1020,1023,1029,1221,1224,1467,1470/Modified site: 4-hydroxyproline (Pro) #status experimental  
 F;627,642,687,708,744,774,795,804,807,810,819,825,846,864,882,897/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F;627,642,687,774,795,804,807,810,819,825,846,864,882,897,1482/Binding site: carbohydrate (Lys) (covalent) #status predicted  
 F;708,744/Binding site: carbohydrate (Lys) (covalent) #status experimental  
 F;1482/Modified site: 5-hydroxylysine (Lys) #status predicted  
 F;1605-1606/Cleavage site: Ala-Asp (procollagen C-endopeptidase) #status predicted  
 F;1639,1645,1662,1671/Disulfide bonds: interchain #status predicted  
 F;1680-1835,1746-1789/Disulfide bonds: #status predicted

Query Match 6.9%; Score 96; DB 1; Length 1838;  
 Best Local Similarity 21.6%; Pred. No. 7.8;  
 Matches 59; Conservative 23; Mismatches 71; Indels 120; Gaps 13;

```

Qy      25 DGPPGSEDPERDDHEGQPRPRVPRKRGHISPKSRPMANSTLLGLLAPPGE----- 74
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1458 DGPPGPMGP-----PGLPGLKGDSPKGE-KGHPGLIGLIGPPGEQGEKGDRLP 1506

Qy      75 -----AWGILGQ-----PPNRPNHSPPPSAKVKKIFGW----- 102
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1507 GPQGSSGPKGEQGITGPSGPPIPPGPPGLPGPPGPKGAKGSSGPTGPKGEAGHPGPPGPP 1566

Qy      103 ---GDFYSNIKTVALNL---LVTGKIVDHNGTFSVHFQHNATGQGNISISLVPPSKAVE 156
      | : : | : : : | | | | | : | |

```

```

Db      1567 GPPGEVIQPLPIQASRTRRNIDASQLLDDGNGENYVDY---ADGM----- 1608
Qy      157 FHQEQQIFIEAKASKIFNCRMWEKVERGRRT-----SLCTH-----DP 195
      :||| : : :| |::| | || || ||
Db      1609 ----EEIF-----GSLNSLKLEIEQMKRPLGTQQNPARTCKDLQLC-HPDFPDGEYWVDP 1658
Qy      196 AKICSRDHAQSSATWSCSQPFKVVCVYIAFYST 228
      : |||| ||| | : | ||
Db      1659 NQGCSRD-----SFKVYCNFTAGGST 1679

```

RESULT 3

S18803  
collagen alpha 1(V) chain - hamster  
C;Species: Cricetinae gen. sp. (hamster)  
C;Date: 19-Mar-1997 #sequence\_revision 24-Jul-1997 #text\_change 16-Dec-1998  
C;Accession: S18803  
R;Greenspan, D.S.; Cheng, W.; Hoffman, G.G.  
J. Biol. Chem. 266, 24727-24733, 1991  
A;Title: The pro-alpha1(V) collagen chain. Complete primary structure, distribution of expression, and comparison with the pro-alpha1(XI) collagen chain.  
A;Reference number: S18802; MUID:92105142; PMID:1722213  
A;Accession: S18803  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-1843 <GRE>  
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology  
F;1620-1842/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 6.8%; Score 94; DB 2; Length 1843;  
Best Local Similarity 24.0%; Pred. No. 12;  
Matches 60; Conservative 24; Mismatches 92; Indels 74; Gaps 10;

```

Qy      25 DGPPGSEDPERDDHEGQPRPRVPRKRGHISPKSRPMANSTLLGLLAPPGE----- 74
      ||||| | :| :| : || : |||: ||||
Db      1463 DGPPGPMGP-----PGLPGLKGDSPGPKGE-KGHPGLIGLIGPPGEQGEKGDRLP 1511
Qy      75 -----AWGILGQ-----PPNRPNHSPPPSAKVKKIFGWGDFYSNIKTVALNLLVT 119
      ||| ||| ||| ||| : | | : :
Db      1512 GPQGSSGPKGEQGITGPSGLPGLPPGPPGLPGLPPGPKGAK---GSSGPTGPKGEAGHPGLP 1568
Qy      120 GKIVDHGNGTFSVHFQHNATGQGNISISLVPPSKAVEFHQEQQIFIEAKASKIFNCRMWEW 179
      | : | : | || | : | | : : :| : : :|
Db      1569 GPPGPPGEVIQPLPIQASRT-RRNIDASQLLDDGAGESYLDYADGMEEIFGSLNSLKLEI 1627
Qy      180 EKVERGRRT-----SLCTH-----DPAKICSRDHAQSSATWSCSQPFKV 218
      |::| | ||| || : |||| |||
Db      1628 EQMKRPLGTQQNPARTCKDLQLC-HPDFPDGEYWVDPNQGCSRD-----SFKV 1674
Qy      219 VCVYIAFYST 228
      | : | ||
Db      1675 YCNFTAGGST 1684

```

RESULT 4

VCVWEK  
env polyprotein - AKV murine leukemia virus  
N;Contains: knob protein gp76; R protein; spike protein p15E  
C;Species: AKV murine leukemia virus  
C;Date: 05-Apr-1983 #sequence\_revision 03-Aug-1984 #text\_change 16-Jul-1999  
C;Accession: A92995; A93448; A03984  
R;Herr, W.  
J. Virol. 49, 471-478, 1984  
A;Title: Nucleotide sequence of AKV murine leukemia virus.  
A;Reference number: A92995; MUID:84115072; PMID:6319746  
A;Accession: A92995  
A;Molecule type: genomic RNA  
A;Residues: 1-669 <HER>  
A;Cross-references: GB:J01998; GB:J01999; GB:K00016; GB:K00017; GB:K00018;  
GB:K01394; NID:g331993; PIDN:AAB03092.1; PID:g331996  
R;Herr, W.; Corbin, V.; Gilbert, W.  
Nucleic Acids Res. 10, 6931-6944, 1982  
A;Title: Nucleotide sequence of the 3' half of AKV.  
A;Reference number: A93448; MUID:83090450; PMID:6294621  
A;Accession: A93448  
A;Molecule type: DNA  
A;Residues: 1-34,'R',36-462,'K',464-591,'K',593-669 <HE2>  
C;Genetics:  
A;Gene: env  
C;Superfamily: type C retrovirus env polyprotein  
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein;  
transmembrane protein  
F;1-31/Domain: signal sequence #status predicted <SIG>  
F;32-470/Product: knob protein gp76 #status predicted <KNB>  
F;471-650/Product: spike protein p15E #status predicted <SPK>  
F;651-669/Product: R protein #status predicted <RPT>  
F;43,199,327,359,399/Binding site: carbohydrate (Asn) (covalent) #status  
predicted

Query Match 6.7%; Score 92.5; DB 1; Length 669;  
Best Local Similarity 23.8%; Pred. No. 4.6;  
Matches 62; Conservative 24; Mismatches 78; Indels 97; Gaps 15;

Qy	24	DDGP--PGSEDPERDDHEGQPRPRVPRKRGHISPKSRPMANSTLLGLLAPPGEAWGILGQ	81
		:                 :	
Db	261	DSGPRVPVIGPNPVLSDRRPPSRPRPTR-----SPPPSNST-----PTET-----	299
Qy	82	PPNRPNHSPPPSAKVKKIFGWGDFYSNIKTV--ALNL-----LVTGKIVDHG-	126
		: : : :           :	
Db	300	PLTLP--EPPPAGVENRLL-----NLVKGAYQALNLTSPDKTQECWLCLVSGPPYYEGV	351
Qy	127	--NGTFSVH-----FQH-----NATGQGNISISLVPPSKAVEFHQEQQIFIEAKA	169
		:                 :     :   :	
Db	352	AVLGTYSNHTSAPANCSVASQHKLTLSSEVTGQG-LCIGAVPKTHQVLCNTTQK-----	403
Qy	170	SKIFNCRMWEKVERGRRTSLCTHDPAKICSRDHAQSSATWSCSQPF-KVVCVYIAFYST	228
		: :   :   :     :	
Db	404	-----TSDGSYYLA---APTGTTWACSTGLTPCISTTILDLT	438
Qy	229	DYRLVQKVCPDYNHSDTPYY	249
		: : :       :	
Db	439	DYCVLVELWPRVTYHSPSYVY	459



# RESULT 5

D40750

proline-rich protein PRB1/2S (EA) - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 03-May-1996

C;Accession: D40750

R;Azen, E.A.; Latreille, P.; Niece, R.L.

Am. J. Hum. Genet. 53, 264-278, 1993

A;Title: PRB1 gene variants coding for length and null polymorphisms among human salivary Ps, PmF, PmS, and Pe proline-rich proteins (PRPs).

A;Reference number: A40750; MUID:93304421; PMID:8317492

A;Accession: D40750

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-117 <AZE>

A;Cross-references: GB:S62930

C;Superfamily: proline-rich protein

Query Match 6.6%; Score 91; DB 2; Length 117;  
Best Local Similarity 31.0%; Pred. No. 0.75;  
Matches 22; Conservative 7; Mismatches 32; Indels 10; Gaps 1;

Qy 22 GQDDGPPGSEDPERDDHEGQPRPRVPRKRGHISPKSRPMANSTLLGLLAPPGEAWGILGQ 81  
|: ||| | | :|: | :| | :| | :|||: | |  
Db 41 GKPQGGPPQGDKSRSPRSPGKPGPPQGGNQPQGGP-----SPPGKPGGPPQ 90

Qy 82 PPNRPNHSPPP 92  
||| |||  
Db 91 GGNRPQGPPPP 101

# RESULT 6

WJHU3E

homeotic protein Hox C4 - human

N;Alternate names: homeotic protein cp19; homeotic protein cp8; homeotic protein Hox 3E

C;Species: Homo sapiens (man)

C;Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 22-Jun-1999

C;Accession: S01030; S15545

R;Simeone, A.; Pannese, M.; Acampora, D.; d'Esposito, M.; Boncinelli, E.

Nucleic Acids Res. 16, 5379-5390, 1988

A;Title: At least three human homeoboxes on chromosome 12 belong to the same transcription unit.

A;Reference number: S01030; MUID:88262550; PMID:2898768

A;Accession: S01030

A;Molecule type: mRNA

A;Residues: 1-264 <SIM>

A;Cross-references: EMBL:X07495; NID:g32385; PIDN:CAA30376.1; PID:g32386

A;Note: the sequence from Fig. 4 is inconsistent with that from Fig. 3 in lacking 108-Ala and 109-Ser and in having 259-Gly

R;Boncinelli, E.; Acampora, D.; Pannese, M.; d'Esposito, M.; Somma, R.; Gaudino, G.; Stornaiuolo, A.; Cafiero, M.; Faiella, A.; Simeone, A.

Genome 31, 745-756, 1989

A;Title: Organization of human class I homeobox genes.

A;Reference number: S15036; MUID:90215256; PMID:2576652

A;Accession: S15545  
A;Molecule type: DNA  
A;Residues: 156-221 <BON>  
C;Genetics:  
A;Gene: GDB:HOXC4  
A;Cross-references: GDB:120672; OMIM:142974  
A;Map position: 12q13.3-12q13.3  
C;Function:  
A;Description: control of embryonic development by tissue- and stage-specific regulation of transcription  
C;Superfamily: homeotic protein Hox D4; homeobox homology  
C;Keywords: alternative splicing; DNA binding; embryo; homeobox; nucleus; transcription regulation  
F;157-213/Domain: homeobox homology <HOX>

Query Match 6.5%; Score 90.5; DB 1; Length 264;  
Best Local Similarity 22.2%; Pred. No. 2.2;  
Matches 47; Conservative 25; Mismatches 91; Indels 49; Gaps 9;

```

Qy      27 PPGSEDPERD-----DHEGQPRPRVPRKRGHISP-KSRPMAN-STLLGLLAPPGEAWGI 78
      ||  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      57 PPRPSYPERQYSCTSLQGPGNSRGGHPAQAGHHHPEKSQSLCEPAPLSGASASPSPAPPA 116

Qy      79 LGQPPNRPNHSPPPSAKVKKIFGWGDFYSNIKTVALNLLVTGKIVDHGNGTFS----- 131
      ||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     117 CSQP--APDHPSSAASKQPIVYPW---MKKIHVSTVNPNYNGGEPKRSRAAYTRQQVLEL 171

Qy     132 ---VHFQHNATGQGNISISLVPPSKAVEFHQEQQIFIEAKASKIF--NCRMEWEKVERGR 186
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     172 EKEFHYNRYLTRRRRIEIA-----HSLCLSERQIKIWFQNRMRMKWKDHRLP 218

Qy     187 RTSLCTHDPAKIC-----SRDHAQSS 207
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     219 NTKVRSAPPAGAAPSTLSAATPGTSEDHSQSA 250

```

#### RESULT 7

JG0183

myosin Myok - Dictyostelium discoideum

C;Species: Dictyostelium discoideum

C;Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 17-May-2002

C;Accession: JG0183

R;Yazu, M.; Adachi, H.; Sutoh, K.

Biochem. Biophys. Res. Commun. 255, 711-716, 1999

A;Title: Novel Dictyostelium unconventional myosin Myok is a class I myosin with the longest loop-1 insert and the shortest tail.

A;Reference number: JG0183; MUID:99160418; PMID:10049776

A;Accession: JG0183

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-858 <YAZ>

A;Cross-references: DDBJ:AB017909

C;Superfamily: myosin heavy chain; myosin motor domain homology

F;10-807/Domain: myosin motor domain homology #status atypical <MMO>

Query Match 6.5%; Score 90.5; DB 2; Length 858;  
Best Local Similarity 23.7%; Pred. No. 9.2;

Matches 46; Conservative 18; Mismatches 77; Indels 53; Gaps 10;

```
Qy      26 GPP---GSEDPERDDHEGQPRPRVPRKRGHISPKSR-----PMANSTLLGLLAPPGEAWG 77
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     176 GPPSRGGGPPPTRG--RGGPPPIQNRGAPPPVSNNGGAPPPVAR----GPVAPPPTR-- 227

Qy      78 ILGQPPNR-----PNHSPPPSAKVKKIFGWGDFYSNIKTV-----AL 114
      |  ||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     228 --GAPPTRGGGPANRGGRGGGPPP--VSTSRGGGGYGGSSKTVDVEHIKKVILDSNPLM 282

Qy     115 NLLVTGKIVDHGNGT-----FSVHFQHNATGQGNISISLVPPSKAVEFHQEQQ----IFI 165
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     283 EAIGNAKTVRNDNSSRFGKYLEIQFDDNNAPVGGLISTFLLKTRVTFQQKNERNFHFIFY 342

Qy     166 EAKASKIFNCRMEW 179
      :  :  :  |  |
Db     343 QMLGGLDQTTKSEW 356
```

RESULT 8

T02885

peroxisome proliferator-activated receptor gamma binding protein PBP165 - mouse

C;Species: Mus musculus (house mouse)

C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 08-Oct-1999

C;Accession: T02885

R;Zhu, Y.; Qi, C.; Jain, S.; Rao, M.S.; Reddy, J.K.

J. Biol. Chem. 272, 25500-25506, 1997

A;Title: Isolation and characterization of PBP, a protein that interacts with peroxisome proliferator-activated receptor.

A;Reference number: Z14760; MUID:97467333; PMID:9325263

A;Accession: T02885

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-1560 <ZHU>

A;Cross-references: EMBL:AF000294; NID:g3411010; PIDN:AAC31118.1; PID:g3411011

Query Match 6.5%; Score 90; DB 2; Length 1560;

Best Local Similarity 22.2%; Pred. No. 21;

Matches 42; Conservative 29; Mismatches 70; Indels 48; Gaps 9;

```
Qy      35 RDDHEGQPRPRVPRKRGHISPKSRPMANSTLLGLLAPPGEAWGILGQPPNRPNHSPPPSA 94
      :|  ||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     568 KDRHESV-----GHGEDFSKVSQNPILTSLLQITGNGGSTIGSSPTPPHHTPPPV 618

Qy      95 KVKKIFGWGDFYSNIKT--VALNLLVTGKIVDHGN--GTFSVHFQHNATGQGNISI---- 146
      :  :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     619 SMA-----GNTKNHPMLMNLKDNPAQDFSTLYGSSPLERQNSSSGSPRMEMCSGS 669

Qy     147 -----SLVPPSKAVEFHQEQQIF-----IEAKASKIFNCRMEWEKVERGRRT-- 188
      |  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     670 NKAKKKKSSRVPPDKPK--HQTEDDFQRELFSMDVDSQ-NPMFDVSMTADALDTPHITPA 726

Qy     189 -SLCTHDPA 196
      |  |  |  |
Db     727 PSQCSTPPA 735
```

# RESULT 9

A46511

envelope protein - AKV murine leukemia virus

C;Species: AKV murine leukemia virus

C;Date: 18-Jun-1993 #sequence\_revision 25-Apr-1997 #text\_change 30-May-1997

C;Accession: A46511

R;Hayashi, H.; Matsubara, H.; Yokota, T.; Kuwabara, I.; Kanno, M.; Koseki, H.; Isono, K.; Asano, T.; Taniguchi, M.

J. Immunol. 149, 1223-1229, 1992

A;Title: Molecular cloning and characterization of the gene encoding mouse melanoma antigen by cDNA library transfection.

A;Reference number: A46511; MUID:92364323; PMID:1380036

A;Accession: A46511

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-669 <HAY>

A;Note: sequence inconsistent with the nucleotide translation

A;Note: sequence extracted from NCBI backbone (NCBIN:110845, NCBIP:110846)

C;Superfamily: type C retrovirus env polyprotein

Query Match 6.5%; Score 89.5; DB 2; Length 669;  
Best Local Similarity 23.8%; Pred. No. 8.3;  
Matches 62; Conservative 23; Mismatches 79; Indels 97; Gaps 15;

```

Qy      24 DDGP--PGSEDPERDDHEGQPRPRVPRKRGHISPKSRPMANSTLLGLLAPPGEAWGILGQ 81
      | | | | : | | | | | | | | | | | | | | | | |
Db      261 DSGPRVPIGPNPVLSDRRPPSRPRPTR-----SPPPSNST-----PTET----- 299

Qy      82 PPNRPNHSPPPSAKVKKIFGWGDFYSNIKTV--ALNL-----LVTGKIVDHG- 126
      | | | | : | | | | | | | | | | | | | | | | |
Db      300 PLTLP--EPPPAGVENRLL-----NLVKGAYQALNLTSPDKTQECWLCLVSGPPYYEGV 351

Qy     127 --NGTFSVH-----FQH-----NATGQGNISISLVPPSKAVEFHQEQQIFIEAKA 169
      | | | | | | | | | | | | | | | | | | | | | |
Db     352 AVLGTYSNHTSAPANCSVASQHKLTLSEVTGQG-LCIGAVPKTHQVLCNTTQK----- 403

Qy     170 SKIFNCRMWEKVERGRRTSLCTHDPAKICSRDHAQSSATWSCSQPF-KVVCVYIAFYST 228
      | | | | : | | | | | | | | | | | | | | | | |
Db     404 -----TSDGSYYLV---APTGTTWACSTGLTPCISTTILNLTT 438

Qy     229 DYRLVQKVC PDYNYHS DTPYY 249
      | | : : : | | | : |
Db     439 DYCVELWPRVTYHSPSYVY 459

```

## RESULT 10

D86245

hypothetical protein [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: D86245

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.; Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.; Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.; Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.

Search completed: June 14, 2004, 19:01:36  
Job time : 62 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2004, 19:00:26 ; Search time 22 Seconds  
(without alignments)  
591.352 Million cell updates/sec

Perfect score: 1386  
Sequence: 1 MQLTRCCFVFLVQGSLLYLV.....VQKVCPDYNHSDTPYPSG 252

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1378	99.4	252	4	US-09-976-594-112	Sequence 112, App
2	92	6.6	1581	3	US-09-110-517-2	Sequence 2, Appli
3	91	6.6	125	4	US-09-252-991A-21090	Sequence 21090, A
4	88.5	6.4	681	4	US-09-252-991A-24567	Sequence 24567, A
5	87.5	6.3	180	3	US-09-187-331-5	Sequence 5, Appli
6	87.5	6.3	180	4	US-09-470-946-5	Sequence 5, Appli
7	86	6.2	371	4	US-09-252-991A-23484	Sequence 23484, A
8	86	6.2	548	4	US-09-252-991A-20793	Sequence 20793, A
9	85.5	6.2	549	1	US-08-494-168-8	Sequence 8, Appli

10	85	6.1	1073	4	US-09-252-991A-27341	Sequence 27341, A
11	84.5	6.1	609	4	US-09-396-149-6	Sequence 6, Appli
12	84	6.1	195	3	US-09-187-331-1	Sequence 1, Appli
13	84	6.1	195	4	US-09-470-946-1	Sequence 1, Appli
14	84	6.1	453	2	US-08-484-126-3	Sequence 3, Appli
15	84	6.1	453	4	US-09-374-909-3	Sequence 3, Appli
16	83	6.0	219	4	US-09-252-991A-17169	Sequence 17169, A
17	82.5	6.0	659	4	US-09-562-737-20	Sequence 20, Appl
18	82	5.9	207	2	US-08-426-599B-2	Sequence 2, Appli
19	82	5.9	207	2	US-08-426-599B-4	Sequence 4, Appli
20	82	5.9	369	4	US-09-252-991A-25394	Sequence 25394, A
21	82	5.9	381	4	US-09-510-031A-5	Sequence 5, Appli
22	82	5.9	770	4	US-09-252-991A-28510	Sequence 28510, A
23	81.5	5.9	457	4	US-09-252-991A-17452	Sequence 17452, A
24	81.5	5.9	822	4	US-09-252-991A-21920	Sequence 21920, A
25	81	5.8	207	3	US-08-813-884-54	Sequence 54, Appl
26	81	5.8	282	4	US-09-252-991A-19978	Sequence 19978, A
27	80.5	5.8	551	4	US-09-252-991A-28829	Sequence 28829, A
28	80	5.8	194	4	US-09-134-000C-6266	Sequence 6266, Ap
29	80	5.8	361	4	US-09-252-991A-24065	Sequence 24065, A
30	80	5.8	400	4	US-09-252-991A-18589	Sequence 18589, A
31	79.5	5.7	400	4	US-09-252-991A-31900	Sequence 31900, A
32	79.5	5.7	480	4	US-09-252-991A-27498	Sequence 27498, A
33	79.5	5.7	596	4	US-09-252-991A-23233	Sequence 23233, A
34	79	5.7	161	4	US-09-489-039A-9917	Sequence 9917, Ap
35	79	5.7	207	4	US-09-224-514A-2	Sequence 2, Appli
36	79	5.7	207	4	US-09-224-514A-10	Sequence 10, Appl
37	79	5.7	207	6	5221624-1	Patent No. 5221624
38	79	5.7	716	1	US-08-396-479B-4	Sequence 4, Appli
39	79	5.7	716	1	US-08-818-823-4	Sequence 4, Appli
40	79	5.7	716	3	US-09-037-190-38	Sequence 38, Appl
41	79	5.7	716	3	US-09-037-192-38	Sequence 38, Appl
42	79	5.7	716	3	US-09-037-143-38	Sequence 38, Appl
43	79	5.7	716	3	US-09-049-691-38	Sequence 38, Appl
44	79	5.7	716	3	US-08-260-174-38	Sequence 38, Appl
45	79	5.7	716	4	US-09-338-128A-38	Sequence 38, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-976-594-112

; Sequence 112, Application US/09976594

; Patent No. 6673549

; GENERAL INFORMATION:

; APPLICANT: Furness, Michael

; APPLICANT: Buchbinder, Jenny

; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS

; FILE REFERENCE: PA-0041 US

; CURRENT APPLICATION NUMBER: US/09/976,594

; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: 60/240,409

; PRIOR FILING DATE: 2000-10-12

; NUMBER OF SEQ ID NOS: 1143

; SOFTWARE: PERL Program

; SEQ ID NO 112  
; LENGTH: 252  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. 6673549 3070147CD1  
US-09-976-594-112

Query Match 99.4%; Score 1378; DB 4; Length 252;  
Best Local Similarity 99.6%; Pred. No. 5.2e-149;  
Matches 251; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 MQLTRCCFVFLVQGSLYLVICGQDDGPPGSEDPERDDHEGQPRPRVPRKRGHISPKSRPM 60
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MQLTRCCFVFLVQGSLYLVICGQDDGPPGSEDPERDDHEGQPRPRVPRKRGHISPKSRPM 60

Qy     61 ANSTLLGLLAPPGEAWGILGQPPNRPNHSPPPSAKVKKIFGWGDFYSNIKTVALNLLVTG 120
      ||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     61 ANSTLLGLLAPTGEAWGILGQPPNRPNHSPPPSAKVKKIFGWGDFYSNIKTVALNLLVTG 120

Qy    121 KIVDHGNGTFSVHFQHNATGQGNISISLVPPSKAVEFHQEQQIFIEAKASKIFNCRMWE 180
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 KIVDHGNGTFSVHFQHNATGQGNISISLVPPSKAVEFHQEQQIFIEAKASKIFNCRMWE 180

Qy    181 KVERGRRTSLCTHDPAKICSRDHAQSSATWSCSQPFKVVVCVYIAFYSTDYRLVQKVC PDY 240
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 KVERGRRTSLCTHDPAKICSRDHAQSSATWSCSQPFKVVVCVYIAFYSTDYRLVQKVC PDY 240

Qy    241 NYHSDTPYYPSG 252
      ||||||||||||
Db    241 NYHSDTPYYPSG 252
```

RESULT 2

US-09-110-517-2

; Sequence 2, Application US/09110517A  
; Patent No. 6248520  
; GENERAL INFORMATION:  
; APPLICANT: Roeder, Robert G  
; APPLICANT: Fondell, Joseph D  
; APPLICANT: Yuan, Chao X  
; APPLICANT: Ito, Mitsuhiro  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING NUCLEAR HORMONE  
; TITLE OF INVENTION: RECEPTOR COACTIVATORS AND USES THEREOF  
; FILE REFERENCE: 600-1-224  
; CURRENT APPLICATION NUMBER: US/09/110,517A  
; CURRENT FILING DATE: 1998-07-06  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1581  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-110-517-2

Query Match 6.6%; Score 92; DB 3; Length 1581;

Best Local Similarity 23.0%; Pred. No. 0.68;  
Matches 43; Conservative 26; Mismatches 72; Indels 46; Gaps 9;

```
Qy      35 RDDHEGQPRPRVPRKRGHISPKSRPMANSTLLGLLAPPGEAWGILGQPPNRPNHSPPPSA 94
      :| ||      ||      |: | | ||      |      :| | | :|: || | :
Db      583 KDRHESV-----GHGEDFSKVSQNPILTSLLQITGNGGSTIGSSPTPPHHTPPPVVS 633

Qy      95 KVKKIFGWGDFYSNIKT--VALNLLVTGKIVDHGN--GTFSVHFQHNATGQGNISI---- 146
      :      | |      : : || |      |      |: : |::: | : |
Db      634 SMA-----GNTKNHPMLMNLKDNPAQDFSTLYGSSPLERQNSSSGSPRMEICSGS 684

Qy      147 -----SLVPPSKAVEFHQEQQIF-----IEAKASK-IFNCRMWEKVERGRRT--- 188
      | : || |      || : |      : : : | ||: | : : : |
Db      685 NKTKKKKSSRLPPEKPK--HQTEDDFQRELFSDVDSDQKPIFDVNMATDTPHITPAP 742

Qy      189 SLCTHDP 195
      | | : |
Db      743 SQCSTPP 749
```

### RESULT 3

US-09-252-991A-21090

; Sequence 21090, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 21090

; LENGTH: 125

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

; FEATURE:

; NAME/KEY: UNSURE

; LOCATION: (79)

; OTHER INFORMATION: Identity of amino acid at the above locations are  
unknown.

US-09-252-991A-21090

Query Match 6.6%; Score 91; DB 4; Length 125;

Best Local Similarity 28.2%; Pred. No. 0.019;

Matches 35; Conservative 19; Mismatches 44; Indels 26; Gaps 8;

```
Qy      33 PERDDHEGQPRPRVP----RKRGHISPKSRPMANSTLLGLLAPPGEAWGILGQPPNRPNH 88
      ||| || ||: | | | : || | | : : | ||
Db      1 PERPDH-AQPQPHFPFCPARRRGH---GRRRAARAVVRARSAAIGE-----PQR 45

Qy      89 SPPPSAKVKKIFGWGDFYSNIKTVALNLLVTGKIVD---HGNGTFS-VHFQHNATGQGNI 144
```



```

      : | : :: | : | | | | : | : | | : : | | :
Db      46 AAPATGIMETIRNYG--YDIIMLVAL-LVVASMFIGXWYHAYGTYAEIHTGRKTWGQFGL 102

Qy      145 SISL 148
      : : :
Db      103 TVAI 106

```

RESULT 4

```

US-09-252-991A-24567
; Sequence 24567, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24567
; LENGTH: 681
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24567

```

```

Query Match          6.4%; Score 88.5; DB 4; Length 681;
Best Local Similarity 26.8%; Pred. No. 0.48;
Matches 33; Conservative 10; Mismatches 31; Indels 49; Gaps 6;

```

```

Qy      22 GQDDGPPGSEDPERDDHEGQPR-----PRVPRKRGHISPKSR-----P 59
      | : | | | | : | | | | | | | : | | : : | |
Db      240 GRGDGPPGEPGADRLDSGGGPAHAPAAPTPRLHRRARQSLAPRRRIPPRSRRGPAAPLP 299

Qy      60 MAN-----STLLGLLAPPGEAWG--IL-----GQP-----PNRPNHSPPP 92
      : : | | | | | | : | | | | | | | |
Db      300 LRDPAGAPGRRRPRLETLAGAAAPRRAGGGRVLRRSRRGSGLPCLRPVADPTLPATCPAP 359

Qy      93 SAK 95
      | :
Db      360 GAR 362

```

RESULT 5

```

US-09-187-331-5
; Sequence 5, Application US/09187331
; Patent No. 6043056
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Gorgone, Gina A.

```

```
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS
; FILE REFERENCE: PF-0631 US
; CURRENT APPLICATION NUMBER: US/09/187,331
; CURRENT FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: g2499136
US-09-187-331-5
```

```
Query Match          6.3%; Score 87.5; DB 3; Length 180;
Best Local Similarity 20.8%; Pred. No. 0.082;
Matches 44; Conservative 26; Mismatches 77; Indels 65; Gaps 9;
```

```
Qy      7 CFVFLVQGSPLYLVICGQDDGPPGSEDPERDDHEG-----QPRPRVPRKRGHISPKSRPM- 60
      ||:  :|      :  ||  ::  |  |  :  ||  |  |  |  |  |  |  |  |  |  |
Db     14 CFLMHARGQRDFDLADALDDPEPTKKPNSDIYPKPKPPYPQPENPDSSGNIYPRPKPRP 73

Qy     61 -----ANSTLLGLLAPPGEAWGIL-----GQPPNRPNHSPPPSAKVKKIFGWGDFYSNI 109
      ||      |  :  |      |  |  ||  ||      |  |  ||:
Db     74 QPQPGNS-----GNSGGYFNDVDRDDGRYP RPRPRPPAG-----GGGGGYSS- 116

Qy    110 KTVALNLLVTGKIVDHGNGTFSVHFQHNAT---GQGNISISLVPPSKAVEFHQEQQIFIE 166
      :||  :  ||:  |  ||:  :|  |  :|  :  :
Db    117 -----YGNSDNTHGGDHHSTYGNPEGNMVAKIVSPIVSVVV-----VTLL 156

Qy    167 AKASKIFNCRMWEKVERGRRTSLCTHDPABI 198
      |:  |      ||  ||:  |  :
Db    157 GAAASYFKL-----NNRRNCFRTHPEENV 180
```

# RESULT 6

```
US-09-470-946-5
; Sequence 5, Application US/09470946
; Patent No. 6358923
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Gorgone, Gina A.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS
; FILE REFERENCE: PF-0631 US
; CURRENT APPLICATION NUMBER: US/09/470,946
; CURRENT FILING DATE: 1999-12-22
; EARLIER APPLICATION NUMBER: US 09/187,331
; EARLIER FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 180
; TYPE: PRT
```

; ORGANISM: Homo sapiens  
; FEATURE: -  
; OTHER INFORMATION: g2499136  
US-09-470-946-5

Query Match 6.3%; Score 87.5; DB 4; Length 180;  
Best Local Similarity 20.8%; Pred. No. 0.082;  
Matches 44; Conservative 26; Mismatches 77; Indels 65; Gaps 9;

```
Qy      7 CFVFLVQGSLYLVICGQDDGPPGSEDPERDDHEG-----QPRPRVPRKRGHISPKSRPM- 60
      ||:  :|  :  :  ||  ::  |  |  :  :|  |  |  :|  ||  :  :|
Db     14 CFLMHARGQRDFDLADALDDPEPTKKPNSDIYPKPKPPYYPQENPDSGGNIYPRPKPRP 73

Qy     61 ----ANSTLLGLLAPPGEAWGIL-----GQPPNRPNHSPPPSAKVKKIFGWGDFYSNI 109
      ||  ||  :  |  :  |  |  |  ||  ||  |  |  ||  :
Db     74 QPQPGNS-----GNSGGYFNDVDRDDGRYP RPRPRPPAG-----GGGGGYSS- 116

Qy    110 KTVALNLLVTGKIVDHGNGTFSVHFQHNAT---GQGNISISLVPPSKAVEFHQEQQIFIE 166
      :||  :  :||  :||  :  |  |  :  :
Db    117 -----YGNSDNTHGGDHHSTYGNPEGNMVAKIVSPIVSVVV-----VTLL 156

Qy    167 AKASKIFNCRMWEKVERGRRTSLCTHDPABI 198
      |:  |  ||  ||:  :
Db    157 GAAASYFKL-----NNRRNCFRTHPEENV 180
```

#### RESULT 7

US-09-252-991A-23484  
; Sequence 23484, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23484  
; LENGTH: 371  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23484

Query Match 6.2%; Score 86; DB 4; Length 371;  
Best Local Similarity 23.3%; Pred. No. 0.37;  
Matches 44; Conservative 19; Mismatches 60; Indels 66; Gaps 9;

```
Qy     25 DGPPGSEDPERDDHEGQPRP---RVPRKRGH-----ISPKSRPMA 61
      |||  |:|  |  ||  |  :||  :  :|
Db      3 DGPDRPPRPDRPGHRVQGRPAYRRSPHRRGHRHHPRPGLRQGHRRQEGHPPLRPRLRAPR 62
```

```

Qy      62 NSTLLG---LL--APPGEA-----WGIL-----GQP-----PNRPNH 88
      : |  ||  | | : |
Db      63 RGAVAGGHRLLRAARPADARSLYPCQRRRLRRRPVHGILPGLRQPRPGDPAHRQPARAQH 122

Qy      89 SPPPSAKVKKI-----FGWGDFYSNIKTVALNLLVTGKIVDHGNGTF-SVHFQHNA 138
      ||  : | : |  |  : :  : : |  : : | |  ||
Db      123 PPPDRDRVQGLRPRAAHGHRAGRADGRADAVDQRPVPMQTVAVIDYGMGNLHSVAKALEH 182

Qy      139 TGQGNISIS 147
      | | : : |
Db      183 VGAGRVLVS 191

```

RESULT 8

```

US-09-252-991A-20793
; Sequence 20793, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20793
; LENGTH: 548
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20793

```

```

Query Match          6.2%; Score 86; DB 4; Length 548;
Best Local Similarity 34.6%; Pred. No. 0.66;
Matches 28; Conservative 6; Mismatches 35; Indels 12; Gaps 5;

```

```

Qy      30 SEDPERDDHE----GQPRPRVPRKRGHISP--KSRPMANSTLL--GLL--APPGEAWGIL 79
      | | : |  | | || ||: |  | : ||  ||  ||  ||  ||
Db      223 SRSPRQQRHPGGTGGDSRPGAPRRRQRADPVVRRRPHPGPALLPRPLLPGGPAAATGGIP 282

Qy      80 GQPPNRPN--HSPPPSAKVKK 98
      ||  ||  | | | : : :
Db      283 RQPDGRPCQWHLPAAPQQRVRR 303

```

RESULT 9

```

US-08-494-168-8
; Sequence 8, Application US/08494168
; Patent No. 5731192
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T.
; APPLICANT: Zhou, Jing
; TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method

```

```

; TITLE OF INVENTION: of Detecting Collagen Deficiency
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/494,168
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/112,465
; FILING DATE: 27-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: SAXE, Bernhard D.
; REGISTRATION NUMBER: 28,665
; REFERENCE/DOCKET NUMBER: 40397/104/BABR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 549 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-494-168-8

```

```

Query Match          6.2%; Score 85.5; DB 1; Length 549;
Best Local Similarity 25.9%; Pred. No. 0.76;
Matches 38; Conservative 11; Mismatches 57; Indels 41; Gaps 6;

```

```

QY      26 GPPGSEDPERDDHEGQPRPRVPRKRGHISPKSRPMANSTLLGLLAPPGEAWGILGQ---- 81
      |||| : |: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      157 GPPGPQGPKGQKGEPEYALPKEERDRYRGEP-----GEPGLVGFQGPGRP-GHVGQMGPV 210

QY      82 -PPNRPNHSPPPSAKVKKIFGWGDFYSNIKTVALNLLVTGKIVDHGNGTFSVHFQHNATG 140
      | || | || | | | | | | | | | | | | | | | | | | | | | | | | |
Db      211 GAPGRPGPPGPPGPK-----GQQGNRGLGFYGVKGEKGDVG 246

QY      141 Q----GNISISLVP--PSKAVEFHQEQ 161
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      247 QPGPNGIPSDTLHPPIIAPTGVTFHPDQ 273

```

```

RESULT 10
US-09-252-991A-27341
; Sequence 27341, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:

```

```
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27341
; LENGTH: 1073
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (803)
; OTHER INFORMATION: Identity of amino acid at the above locations are
unknown.
US-09-252-991A-27341
```

```
Query Match          6.1%; Score 85; DB 4; Length 1073;
Best Local Similarity 28.9%; Pred. No. 2.4;
Matches 28; Conservative 6; Mismatches 43; Indels 20; Gaps 3;
```

```
Qy      22 GQDDGPPGSED--PERDDHEGQPRPRVPR-----KRGHISPKSRPMANSTLLG 67
      | : | || | ||| | | | | | | | | | | | | | | | | | | | | | | | |
Db      524 GAEARPAGSSDRPPERDVAAADPHPRGTGRYRGPGAAREHRGNRGGSCPRRSPVSAG---- 579

Qy      68 LLAPPGEAWGILGQPPNRPNHSPPPSAKVKKIFGWGD 104
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      580 --APPAPHPNIQENPNGRQQEQGPAVARCGALHGLGD 614
```

# RESULT 11

```
US-09-396-149-6
; Sequence 6, Application US/09396149
; Patent No. 6538176
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; TITLE OF INVENTION: Maize Replication Protein A and Use
; FILE REFERENCE: 5718-59
; CURRENT APPLICATION NUMBER: US/09/396,149
; CURRENT FILING DATE: 1999-09-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 609
; TYPE: PRT
; ORGANISM: Xenopus laevis
US-09-396-149-6
```

```
Query Match          6.1%; Score 84.5; DB 4; Length 609;
Best Local Similarity 23.2%; Pred. No. 1.2;
Matches 43; Conservative 26; Mismatches 65; Indels 51; Gaps 10;
```

Qy 26 GPPGSEDPERDDHEGQPRPRVPRKRGHISPKSRPMANSTLLGLLAPPGEAWGILGQPPNR 85  
 | | : | | | | : | | : | :  
 Db 110 GKIGNPQPYND---GQPQPAAP-----APASAPAPAPSKL-----Q 142

Qy 86 PNHSPPPSAK--VKKIFGWGDFY-----SNIKTVALNLL-----VTGKIVDHGNGTF 130  
 | : | | | | | : | | | : | | : : | :  
 Db 143 NNSAPPPSMNRGTSKLFGGGSLNTPGGSQSKVVPPIASLNPHYQSKWTVRARVTNKG---- 198

Qy 131 SVHFQHNATGQGNISISLVPPS---KAVEFH-QEQQIFIEAKASKIFNCRMWEKVERG 185  
 : | : | : | : | : | : | : | : : : | :  
 Db 199 QIRTSNSRGEGLFSIEMVDESGEIRATAFNEQADKFFSIEVNVVYFYSKGTCLKIANK 258

Qy 186 RRTSL 190  
 : | :  
 Db 259 QYTSV 263

RESULT 12

US-09-187-331-1

; Sequence 1, Application US/09187331  
 ; Patent No. 6043056  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yue, Henry  
 ; APPLICANT: Corley, Neil C.  
 ; APPLICANT: Guegler, Karl J.  
 ; APPLICANT: Gorgone, Gina A.  
 ; APPLICANT: Baughn, Mariah R.  
 ; TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS  
 ; FILE REFERENCE: PF-0631 US  
 ; CURRENT APPLICATION NUMBER: US/09/187,331  
 ; CURRENT FILING DATE: 1998-11-06  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 1  
 ; LENGTH: 195  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE: -  
 ; OTHER INFORMATION: 2297891  
 US-09-187-331-1

Query Match 6.1%; Score 84; DB 3; Length 195;  
 Best Local Similarity 20.3%; Pred. No. 0.23;  
 Matches 45; Conservative 27; Mismatches 80; Indels 70; Gaps 11;

Qy 7 CFVFLVQGSLYLVICGQDDGPPGSEDPERDDHEG-----QPRPRVPRKRGHISPKSRPM- 60  
 || : : | : | : : | : | : | : : | :  
 Db 14 CFLMHARGQRDFDLADALDDPEPTKKPNSDIYPKPKPPYYPQENPDSSGNIYPRPKPRP 73

Qy 61 -----ANSTLLGLLAPPGEAWGIL-----GQPPNRPNHSPPPSAKVKKIFGWGDF--YS 107  
 || : | : | : | : | : | : | : | :  
 Db 74 QPQPGNS-----GNSGGYFNDVDRDDGRYP RPRPRPPAGG-----GGGGYSSYG 118

Qy 108 NIKTVALNLLVTGKIVDHGNGTFSVHFQ-----HNAT---GQGNISISLVPPPSKAVE 156  
 | : | : : : : | : : | : : | : | :  
 Db 119 NSDNT-----HGRGGYRPNSRYGNTYGGDHHSTYGNPEGNMVAKIVSPIVSVV 166

Qy 157 FHQEQQIFIEAKASKIFNCRMWEKVERGRRTSLCTHDPAKI 198  
 : : | : | || ||:| :  
 Db 167 V-----VTLLGAAASYFKL-----NNRRNCFRTHEPENV 195

RESULT 13

US-09-470-946-1  
 ; Sequence 1, Application US/09470946  
 ; Patent No. 6358923  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yue, Henry  
 ; APPLICANT: Corley, Neil C.  
 ; APPLICANT: Guegler, Karl J.  
 ; APPLICANT: Gorgone, Gina A.  
 ; APPLICANT: Baughn, Mariah R.  
 ; TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS  
 ; FILE REFERENCE: PF-0631 US  
 ; CURRENT APPLICATION NUMBER: US/09/470,946  
 ; CURRENT FILING DATE: 1999-12-22  
 ; EARLIER APPLICATION NUMBER: US 09/187,331  
 ; EARLIER FILING DATE: 1998-11-06  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 1  
 ; LENGTH: 195  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE: -  
 ; OTHER INFORMATION: 2297891  
 US-09-470-946-1

Query Match 6.1%; Score 84; DB 4; Length 195;  
 Best Local Similarity 20.3%; Pred. No. 0.23;  
 Matches 45; Conservative 27; Mismatches 80; Indels 70; Gaps 11;

Qy 7 CFVFLVQGSlyLVICGQDDGPPGSEDPERDDHEG-----QPRPRVPRKRGHISPKSRPM- 60  
 ||: :| : | | :: | | : | : | | :| :| :|  
 Db 14 CFLMHARGQRDFDLADALDDPEPTKKPNSDIYPKPKPPYPQPENPDSSGGNIYPRPKPRP 73  
 Qy 61 ----ANSTLLGLLAPPGEAWGIL-----GQPPNRPNHSPPPSAKVKKIFGWGDF--YS 107  
 || | : | | : | || || | | : |  
 Db 74 QPQPGNS-----GNSGGYFNDVDRDDGRYP RP RP RP PAGG-----GGGGYSSYG 118  
 Qy 108 NIKTVALNLLVTGKIVDHGNGTFSVHFQ-----HNAT---GQGNISISLVPPSKAVE 156  
 | || : : : | :| :||: :| | :|  
 Db 119 NSDNT-----HGRGGYRPNSRYGNTYGGDHHSTYGNPEGNMVAKIVSPIVSVV 166  
 Qy 157 FHQEQQIFIEAKASKIFNCRMWEKVERGRRTSLCTHDPAKI 198  
 : : | : | || ||:| :  
 Db 167 V-----VTLLGAAASYFKL-----NNRRNCFRTHEPENV 195

RESULT 14

US-08-484-126-3  
 ; Sequence 3, Application US/08484126  
 ; Patent No. 5985655



```

; GENERAL INFORMATION:
;   APPLICANT: Anderson, W. French
;   APPLICANT: Baltrucki, Leon F.
;   APPLICANT: Mason, James M.
;   TITLE OF INVENTION: Targetable Vector Particles
;   NUMBER OF SEQUENCES: 8
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
;     ADDRESSEE: Cecchi, Stewart & Olstein
;     STREET: 6 Becker Farm Road
;     CITY: Roseland
;     STATE: New Jersey
;     COUNTRY: USA
;     ZIP: 07068
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: 3.5 inch diskette
;     COMPUTER: IBM PS/2
;     OPERATING SYSTEM: MS-DOS
;     SOFTWARE: Word Perfect 5.1
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/484,126
;     FILING DATE: 07-JUN-1995
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/326,347
;     FILING DATE: 20-OCT-1994
;     APPLICATION NUMBER: 08/973,307
;     FILING DATE: 09-NOV-1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Lillie, Raymond J.
;     REGISTRATION NUMBER: 31,778
;     REFERENCE/DOCKET NUMBER: 271010-281
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 201-994-1700
;     TELEFAX: 201-994-1744
;   INFORMATION FOR SEQ ID NO: 3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 453 amino acids
;       TYPE: amino acid
;       STRANDEDNESS:
;       TOPOLOGY: linear
;     MOLECULE TYPE: protein
;     FEATURE:
;       NAME/KEY: xenotropic gp70 protein
US-08-484-126-3

```

```

Query Match          6.1%; Score 84; DB 2; Length 453;
Best Local Similarity 23.4%; Pred. No. 0.84;
Matches 60; Conservative 25; Mismatches 69; Indels 102; Gaps 16;

```

```

Qy      44 PRVPRKRGHISPKSRPMANSTLLGLLAPPGGEAWGILGQPPNRPNH-----SPPPSAK 95
      |||| | | | | | : | | | : | | | : |||| |
Db      243 PRVP-----IGP-----NPVITDQLPPSQPVQIMLPRPPHPPPSGTVSMVPGAPPPSQQ 291

Qy      96 VKKIFGWGDFYSNI---KTVALNL-----LVTGKIVDHG---NGTFSVH---- 133
      || | | : |||| | | | | | | | | | | | |
Db      292 P----GTGDRLLNLVEGAYQALNLTSPDKTQECWLCLVSGPPYYEGVAVLGTYSNHTSAP 347

```

Qy 134 -----FQH-----NATGQGNISISLVPPSKAVEFHQEQQIFIEAKASKIFNCRMWEK 182  
 || |||| : : || : ||  
 Db 348 ANCSVASQHKLTLEVTGQG-LCVGAVPKT-----HQ----- 378  
 Qy 183 ERGRRTSLC-----THDPAKICSRDHAQSSATWSCSQPF-KVVCVYIAFYSTDYRLVQKV 236  
 :|| | | : : | : | : : : : ||| : : :  
 Db 379 -----ALCNTTQKTS DGSYYLA---APAGTIWACNTGLTPCLSTTVLNLTTDYCVLVEL 429  
 Qy 237 CPDYNVHSDTPYYPG 252  
 | ||| | |  
 Db 430 WPKVTYHS--PDYVYG 443

RESULT 15

US-09-374-909-3

; Sequence 3, Application US/09374909

; Patent No. 6503501

; GENERAL INFORMATION:

; APPLICANT: Anderson, W. French

; Baltrucki, Leon F.

; Mason, James M.

; TITLE OF INVENTION: Targetable Vector Particles

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,

; Cecchi, Stewart & Olstein

; STREET: 6 Becker Farm Road

; CITY: Roseland

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07068

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch diskette

; COMPUTER: IBM PS/2

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Word Perfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/374,909

; FILING DATE: 13-Aug-1999

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/484,126

; FILING DATE: <Unknown>

; APPLICATION NUMBER: 08/973,307

; FILING DATE: 09-NOV-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Lillie, Raymond J.

; REGISTRATION NUMBER: 31,778

; REFERENCE/DOCKET NUMBER: 271010-281

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201-994-1700

; TELEFAX: 201-994-1744

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 453 amino acids

; TYPE: amino acid

Perfect score: 1386  
Sequence: 1 MQLTRCCFVFLVQGSLYLVI.....VQKVCPDYNHSDTPYYPSG 252  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 6019581 seqs, 976053577 residues  
Total number of hits satisfying chosen parameters: 6019581  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : Pending\_Patents\_AA\_Main:\*

Perfect score: 1386  
Sequence: 1 MQLTRCCFVFLVQGSLYLVI.....VQKVCPDYNHSDTPYYPSG 252  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 288743 seqs, 43614698 residues  
Total number of hits satisfying chosen parameters: 288743  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : Pending\_Patents\_AA\_New:\*

```

Qy      7 CFVFLVQGSlyLVICGQDDGPPGSEDPERDDHEG-----QPRPRVPRKRGHISPKSRPM- 60
      ||:  :|  :  |  |  ::  |  |  :  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     14 CFLMHARGQRDFDLADALDDPEPTKKPNSDIYPKPKPPYYPQENPDSSGGNIYPRPKPRP 73

Qy     61 ----ANSTLLGLLAPPGEAWGIL-----GQPPNRPNHSPPPSAKVKKIFGWGDFYSNI 109
      ||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     74 QPQPGNS-----GNSGGYFNDVDRDDGRYPPrPRPRPPAG-----GGGGGYSS- 116

Qy    110 KTVALNLLVTGKIVDHGNGTFSVHFQHNAT---GQGNISISLVPPSKAVEFHQEQQIFIE 166
      :||  :  |  ::|  :||:  :|  |  :|  :  :
Db    117 -----YGNSDNTHGGDHHSTYGNPEGNMVAKIVSPIVSVVV-----VTLL 156

Qy    167 AKASKIFNCRMEWEKVERGRRTSLCTHDPAKI 198
      |:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    157 GAAASYFKL-----NNRRNCFRTHPENV 180

```

Search completed: June 14, 2004, 19:03:57  
Job time : 41 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2004, 18:55:36 ; Search time 18 Seconds  
(without alignments)  
728.982 Million cell updates/sec

Title: US-10-063-599-92  
Perfect score: 1386  
Sequence: 1 MQLTRCCFVFLVQGSlyLVl.....VQKVCpDYNyHSDTPYyPSG 252

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result %  
Query

No.	Score	Match	Length	DB	ID	Description
1	1386	100.0	252	1	NXP3_HUMAN	O95157 homo sapien
2	1341	96.8	252	1	NXP3_MOUSE	Q91vx5 mus musculu
3	1337	96.5	252	1	NXP3_RAT	Q9z2n5 rattus norv
4	661	47.7	271	1	NXP1_HUMAN	P58417 homo sapien
5	655	47.3	271	1	NXP1_RAT	Q63366 rattus norv
6	653.5	47.2	253	1	NXP1_MOUSE	Q61200 mus musculu
7	640.5	46.2	262	1	NXP2_HUMAN	O95156 homo sapien
8	636.5	45.9	264	1	NXP2_BOVIN	Q28145 bos taurus
9	602	43.4	244	1	NXP2_MOUSE	Q61199 mus musculu
10	512	36.9	304	1	NXP4_RAT	Q9z2n4 rattus norv
11	497	35.9	95	1	NXP3_MACMU	Q8wmj4 macaca mula
12	495	35.7	308	1	NXP4_HUMAN	O95158 homo sapien
13	398	28.7	104	1	NXP1_MACMU	Q8wmi6 macaca mula
14	260	18.8	77	1	NXP4_MACMU	Q8wmj7 macaca mula
15	97.5	7.0	264	1	HXC4_MOUSE	Q08624 mus musculu
16	96	6.9	1838	1	CA15_HUMAN	P20908 homo sapien
17	95	6.9	1914	1	RLF_HUMAN	Q13129 homo sapien
18	93.5	6.7	1263	1	BAT8_MOUSE	Q9z148 mus musculu
19	93	6.7	1355	1	CA21_RANCA	O42350 rana catesb
20	92.5	6.7	669	1	ENV_MLVAV	P03386 akv murine
21	91.5	6.6	733	1	HIC1_HUMAN	Q14526 homo sapien
22	90.5	6.5	264	1	HXC4_HUMAN	P09017 homo sapien
23	90.5	6.5	1745	1	CA35_HUMAN	P25940 homo sapien
24	90	6.5	1581	1	PPRB_HUMAN	Q15648 h peroxisom
25	90	6.5	5085	1	PCLO_RAT	Q9jks6 rattus norv
26	89.5	6.5	1210	1	BAT8_HUMAN	Q96kq7 homo sapien
27	89	6.4	2090	1	N214_HUMAN	P35658 homo sapien
28	88.5	6.4	615	1	HIC2_HUMAN	Q96jb3 homo sapien
29	88.5	6.4	1175	1	DSRA_RAT	P55266 rattus norv
30	88	6.3	1286	1	SMC4_MOUSE	Q8cg47 mus musculu
31	87.5	6.3	180	1	XG_HUMAN	P55808 homo sapien
32	87	6.3	675	1	ENV_MLVF5	P03390 friend muri
33	87	6.3	1690	1	CA44_HUMAN	P53420 homo sapien
34	86.5	6.2	1603	1	CA1F_HUMAN	Q07092 homo sapien
35	86	6.2	331	1	PRP1_HUMAN	P04280 homo sapien
36	85.5	6.2	1003	1	MBD6_HUMAN	Q96dn6 homo sapien
37	85.5	6.2	1712	1	CA24_HUMAN	P08572 homo sapien
38	85	6.1	2715	1	MLL4_HUMAN	Q9umn6 homo sapien
39	84.5	6.1	407	1	EVX1_HUMAN	P49640 homo sapien
40	84.5	6.1	463	1	EF12_HUMAN	Q05639 homo sapien
41	84.5	6.1	463	1	EF12_MOUSE	P27706 mus musculu
42	84.5	6.1	609	1	RFA1_XENLA	Q01588 xenopus lae
43	84.5	6.1	743	1	CA18_MOUSE	Q00780 mus musculu
44	84	6.1	998	1	YU55_STRAW	Q82is8 streptomyce
45	83.5	6.0	226	1	PRR2_TOBAC	P07052 nicotiana t

#### ALIGNMENTS

##### RESULT 1

NXP3\_HUMAN

ID NXP3\_HUMAN STANDARD; PRT; 252 AA.

AC O95157; Q8TBF6; Q9ULR1;

DT 16-OCT-2001 (Rel. 40, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Neurexophilin 3 precursor.  
 GN NXP3 OR NPH3 OR KIAA1159.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Hypothalamus;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE OF 32-252 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=20039618; PubMed=10574461;  
 RA Hirose M., Nagase T., Ishikawa K.-I., Kikuno R., Nomura N.,  
 RA Ohara O.;  
 RT "Characterization of cDNA clones selected by the GeneMark analysis  
 RT from size-fractionated cDNA libraries from human brain.";  
 RL DNA Res. 6:329-336(1999).  
 RN [3]  
 RP SEQUENCE OF 80-252 FROM N.A.  
 RX MEDLINE=98237742; PubMed=9570794;  
 RA Missler M., Suedhof T.C.;  
 RT "Neurexophilins form a conserved family of neuropeptide-like  
 RT glycoproteins.";  
 RL J. Mol. Neurosci. 18:3630-3638(1998).  
 CC -!- FUNCTION: May be signaling molecules that resemble neuropeptides.  
 CC Ligand for alpha-neurexins (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted (Potential).  
 CC -!- TISSUE SPECIFICITY: Highest level in brain.  
 CC -!- PTM: May be proteolytically processed at the boundary between the  
 CC N-terminal nonconserved and the central conserved domain in  
 CC neuron-like cells (By similarity).  
 CC -!- SIMILARITY: Belongs to the neurexophilin family.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; BC022541; AAH22541.1; -.  
 DR EMBL; AB032985; BAA86473.1; -.  
 DR EMBL; AF043468; AAD02281.1; -.  
 DR Genew; HGNC:8077; NXPH3.  
 DR MIM; 604636; -.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0007218; P:neuropeptide signaling pathway; NAS.  
 KW Signal; Glycoprotein.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 252 NEUREXOPHILIN 3.  
 FT DOMAIN 23 75 II.  
 FT DOMAIN 76 157 III.  
 FT DOMAIN 158 166 IV (LINKER DOMAIN).  
 FT DOMAIN 167 252 V (CYS-RICH).  
 FT CARBOHYD 62 62 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 137 137 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 143 143 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 178 178 E -> G (IN REF. 1).  
 SQ SEQUENCE 252 AA; 28127 MW; 74D2B3D5A89D221F CRC64;

Query Match 100.0%; Score 1386; DB 1; Length 252;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-107;  
 Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1 MQLTRCCFVFLVQGSLYLVICGQDDGPPGSEDPERDDHEGQPRPRVPRKRGHISPKSRPM 60
      |||
Db      1 MQLTRCCFVFLVQGSLYLVICGQDDGPPGSEDPERDDHEGQPRPRVPRKRGHISPKSRPM 60

Qy     61 ANSTLLGLLAPPGEAWGILGQPPNRPNHSPPPSAKVKKIFGWGDFYSNIKTVALNLLVTG 120
      |||
Db     61 ANSTLLGLLAPPGEAWGILGQPPNRPNHSPPPSAKVKKIFGWGDFYSNIKTVALNLLVTG 120

Qy    121 KIVDHGNGTFSVHFQHNATGQGNISISLVPPSKAVEFHQEQQIFIEAKASKIFNCRMWE 180
      |||
Db    121 KIVDHGNGTFSVHFQHNATGQGNISISLVPPSKAVEFHQEQQIFIEAKASKIFNCRMWE 180

Qy    181 KVERGRRTSLCTHDPAKICSRDHAQSSATWSCSQPFKVVCVYIAFYSTDYRLVQKVC PDY 240
      |||
Db    181 KVERGRRTSLCTHDPAKICSRDHAQSSATWSCSQPFKVVCVYIAFYSTDYRLVQKVC PDY 240

Qy    241 NYHSDTPYYPSG 252
      |||
Db    241 NYHSDTPYYPSG 252
  
```

RESULT 2

NXP3\_MOUSE

ID NXP3\_MOUSE STANDARD; PRT; 252 AA.



AC Q91VX5;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Neurexophilin 3 precursor.  
 GN NXPH3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP TISSUE SPECIFICITY.  
 RX MEDLINE=98237742; PubMed=9570794;  
 RA Missler M., Suedhof T.C.;  
 RT "Neurexophilins form a conserved family of neuropeptide-like  
 RT glycoproteins.";  
 RL J. Neurosci. 18:3630-3638(1998).  
 CC -!- FUNCTION: May be signaling molecules that resemble neuropeptides.  
 CC Ligand for alpha-neurexins.  
 CC -!- SUBCELLULAR LOCATION: Secreted (Potential).  
 CC -!- TISSUE SPECIFICITY: Highest level in brain, present also in lung,  
 CC kidney and testis.  
 CC -!- PTM: May be proteolytically processed at the boundary between the  
 CC N-terminal nonconserved and the central conserved domain in  
 CC neuron-like cells (By similarity).  
 CC -!- SIMILARITY: Belongs to the neurexophilin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; BC007167; AAH07167.1; -.

DR MGD; MGI:1336188; Nxph3.

DR GO; GO:0005102; F:receptor binding; IDA.

KW Signal; Glycoprotein.

FT SIGNAL 1 22 POTENTIAL.

FT CHAIN 23 252 NEUREXOPHILIN 3.

FT DOMAIN 23 75 II.

FT DOMAIN 76 157 III.

FT DOMAIN 158 166 IV (LINKER DOMAIN).

FT DOMAIN 167 252 V (CYS-RICH).

FT CARBOHYD 62 62 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 137 137 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 143 143 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 252 AA; 28183 MW; A770645706435A7C CRC64;

Query Match 96.8%; Score 1341; DB 1; Length 252;  
 Best Local Similarity 96.4%; Pred. No. 1.8e-103;  
 Matches 243; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MQLTRCCFVFLVQGSLLYLVICGQDDGPPGSEDPERDDHEGQPRPRVPRKRGHISPKSRPM 60  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:

Db 1 MQLTRCCFVFLVQGSLLYLVICGQDDGPPGSEDPEHDDHEGQPRPRVPRKRGHISPKSRPL 60

Qy 61 ANSTLLGLLAPPGEAWGILGQPPNRPNHSPPPSAKVKKIFGWGDFYSNIKTVALNLLVTG 120  
 |||||||||||||||||:|||||||||||

Db 61 ANSTLLGLLAPPGEVWGLGQPPNRPKQSPLPSTKVKKIFGWGDFYSNIKTVALNLLVTG 120

Qy 121 KIVDHGNGTFSVHFQHNATGQGNISISLVPPSKAVEFHQEQQIFIEAKASKIFNCRMWE 180  
 |||||||||||||||||:||||||||||||

Db 121 KIVDHGNGTFSVHFRHNATGQGNISISLVPPSKAVEFHQEQQIFIEAKASKIFNCRMWE 180

Qy 181 KVERGRRTSLCTHDPAKICSRDHAQSSATWSCSQPFKVVCVYIAFYSTDYRLVQKVC PDY 240  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 181 KVERGRRTSLCTHDPAKICSRDHAQSSATWSCSQPFKVVCVYIAFYSTDYRLVQKVC PDY 240

Qy 241 NYHSDTPYYPSG 252  
 |||||||||||||

Db 241 NYHSDTPYYPSG 252

# RESULT 3

NXP3\_RAT

ID NXP3\_RAT STANDARD; PRT; 252 AA.

AC Q9Z2N5;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Neurexophilin 3 precursor.

GN NXPH3 OR NPH3.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=98237742; PubMed=9570794;  
 RA Missler M., Suedhof T.C.;  
 RT "Neurexophilins form a conserved family of neuropeptide-like  
 RT glycoproteins.";  
 RL J. Mol. Neurosci. 18:3630-3638(1998).  
 CC -!- FUNCTION: May be signaling molecules that resemble neuropeptides.  
 CC Ligand for alpha-neurexins (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted (Potential).  
 CC -!- TISSUE SPECIFICITY: Brain. Detected in several other tissues.  
 CC -!- PTM: May be proteolytically processed at the boundary between the  
 CC N-terminal nonconserved and the central conserved domain in  
 CC neuron-like cells (By similarity).  
 CC -!- SIMILARITY: Belongs to the neurexophilin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF042713; AAD02226.1; -.  
 KW Glycoprotein; Signal.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 252 NEUREXOPHILIN 3.  
 FT DOMAIN 23 75 II.  
 FT DOMAIN 76 157 III.  
 FT DOMAIN 158 166 IV (LINKER DOMAIN).  
 FT DOMAIN 167 252 V (CYS-RICH).  
 FT CARBOHYD 62 62 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 137 137 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 143 143 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 252 AA; 28225 MW; 5FF10603CE66D8BA CRC64;

Query Match 96.5%; Score 1337; DB 1; Length 252;  
 Best Local Similarity 96.0%; Pred. No. 3.8e-103;  
 Matches 242; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MQLTRCCFVFLVQGSlyLVICGQDDGPPGSEDPERDDHEGQPRPRVPRKRGHISPKSRPM 60  
 |||||:|||||  
 Db 1 MQLTRCCFVFLVQGSlyLVICGQEDGPPGSEDPEHDDHEGQPRPRVPRKRGHISPKSRPL 60  
 Qy 61 ANSTLLGLLAPPGEAWGILGQPPNRPNHSPPPSAKVKKIFGWGDFYSNIKTVALNLLVTG 120  
 |||||:|||||  
 Db 61 ANSTLLGLLAPPGEVWGILGQPPNRPKQSPLPSTKVKKIFGWGDFYSNIKTVALNLLVTG 120  
 Qy 121 KIVDHGNGTFSVHFQHNATGQGNISISLVPPSKAVEFHQEQQIFIEAKASKIFNCRMWE 180  
 |||||:|||||  
 Db 121 KIVDHGNGTFSVHFRHNATGQGNISISLVPPSKAVEFHQEQQIFIEAKASKIFNCRMWE 180  
 Qy 181 KVERGRRTSLCTHDPAKICSRDHAQSSATWSCSQPFKVVVYIAFYSTDYRLVQKVC PDY 240  
 |||||:|||||  
 Db 181 KVERGRRTSLCTHDPAKICSRDHAQSSATWSCSQPFKIVVYIAFYSTDYRLVQKVC PDY 240